International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Second Eye for the Visually Impaired

Nagashree S[1], Sadhana Janardhana[1], Harshith R[1], Prajakta Madankar[2]

[1]Dept. of ISE, The National Institute of Engineering, Mysore
[2]Asst. Professor, Dept. Of ISE, The National Institute of Engineering, Mysore

---

Abstract - *We are proposing this paper called the second eye for the visually impaired, which is devised to detect the text and through the image of the text captured, it produces a voice output. This is achieved by a smart spec which plays a major role.*

A camera will be attached to the spec so that it captures the image of the printed form of the text. A Tesseract-Optical Character Recognition (OCR) will analyze the image and by using a compact open source software speech synthesizer, eSpeak, the text will be converted into speech. TTS method helps to get the voice output of the synthesized speech through a headset.

This paper also proposes wearable equipment that consists of ultrasonic sensors to detect obstacles and their distance which will help avoid any accidents possible due to obstacles that may be encountered on the way.

For the implementation, Raspberry Pi is the main component as it is used to provide an interface between camera, sensors, and image processing results, while also performing functions to manipulate peripheral units (Keyboard, USB etc.,).

Keywords*: *Tesseract OCR (optical Character Recognition), eSpeak, TTS (Text-to-speech), Raspberry Pi, Ultrasonic sensor, Obstacle detection.

1. INTRODUCTION

The recent survey report from WHO confirms that around 285 million people around the world are considered to be visually impaired.Out of all the visually impaired people in the world, around 80% of them have cataract as the cause for their blindness. Rest of the 20% can restore their vision by using eye glasses or contact lenses.

Even though the technology is advanced in many ways, textual information is still the most common mode of information exchange. It is vital to access textual documents in certain situations like reading a text on the go and in some less than ideal conditions. But, accessing textual documents is nearly impossible for the visually impaired people. Therefore, recent technological developments in portable computers, digital cameras and computer vision make it viable to assist these people by developing camera-based products that combine computer vision technology with other existing commercial products like OCR systems.

The smart spec allows the blind people to touch the printed text and receive a speech output in real-time. For this to be implemented, there is a requirement of two technologies, one is OCR (Optical Character Recognition) for Textual Information Extraction (TIE) and the other one is Text-To-Speech Synthesizer (TTS) to convert the extracted text to speech.

In education and employment fields, it is found that the visually impaired people are facing a lot of difficulties in reading and learning contents from book, as it requires flawless reading aid to achieve good academic performance when placed in mainstream or regular institutions. To improve these difficulties in the society, we have proposed this product in the form of spectacles with an in-built camera. This will help them read the captured image of the text and give an audio output.

Normal people even without any visual impairment can also use this product to read a lengthy document in short span of time.

Visually impaired people also face the problems of mobility in an unknown environment. Many efforts have been made to improve their mobility by use of technology. Many people suffer from serious visual impairments preventing them from traveling independently. Accordingly, they need to use a wide range of tools and techniques to help them in their mobility.

Recently, many techniques have been developed to enhance the mobility of visually impaired people that rely on signal processing and sensor technology. These are called Electronic Travel Aid (ETA) devices. ETA helps these people to move freely in an environment regardless of its dynamic changes. This project describes an ETA which is composed of ultrasonic sensors.

Ultrasonic sensors generate high frequency sound waves and evaluate the echo which is received back by the sensor. The distance between the person and the obstacles is calculated by measuring time of the wave travel. The information about the presence of an object is given via vibrating motor.

2. OBJECTIVES

- To read the printed text and receive a speech output in real time.
- To detect obstacles early and inform the user through vibrations.

3. PROPOSED SYSTEM

This paper has a prototype system of the assistive text reading. The concept of the proposed system is to develop specs for text reading mainly for the visually impaired people. There are four main modules in this system, they are, Camera module, Optical Character Recognition module, Text to speech module and Ultrasonic module.

The figure below shows the text reading system and obstacle detection system for the visually impaired users.

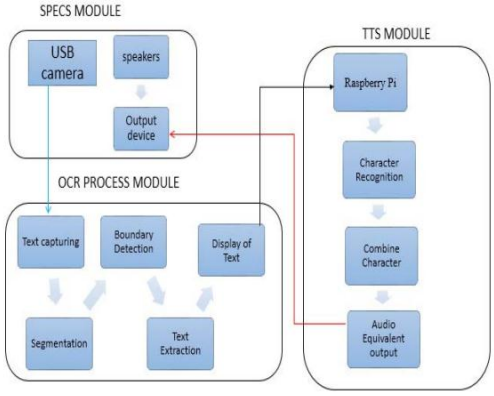

Fig -1: Block Diagram of text detection and conversion

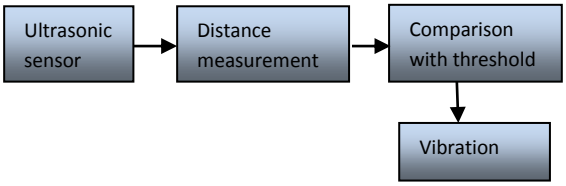

Fig-2: Block Diagram of Obstacle Detection

3.1 Capturing Images

A Raspberry Pi is used for capturing a sequence of images swiftly by utilizing it's video capturing port with JPEG encoder. However there are several issues that need to be addressed, they are:

The video-port capturing is done only when the video is recording. This means that the images captured may not be in the desired resolution or size all the time. i.e., they may be distorted, rotated or blurred etc.

The JPEG encoded images do not have all the required information like co-ordinates, time and they are not exchangeable.

The images captured using video-port are usually "more fragmented" than the ones captured using still-port capture. So, before going through the pre-processed images we have to apply more de-noising algorithms.

All the capture methods found in OpenCV like capture, capture continuous and capture sequence have to be considered according to their use and abilities. Capture sequence method is used in this project as it one of the fastest methods so far. Raspberry Pi camera can capture images at the rate of 20fps and 640×480 resolution using the capture sequence method. While capturing images rapidly, one of the major issues with Raspberry Pi is, Bandwidth. The I/O bandwidth of the Raspberry Pi is limited, because of this the format in which we are pulling pictures makes the process less efficient. Cache exhaustion is also one of the issues, this happens when the SD card size is small and it cannot hold all the images captured by the camera port.

3.2 Capturing and Processing Images

As the Raspberry Pi has limited I/O bandwidth, structuring and multi-threading is an important initial step of the pre-processing algorithms for image. First we need to capture image from video-port and then process it. Raspberry Pi maintains a queue of images and process them as the captured images come in. One very important factor is that the encoder should not be stalled, for this to happen, the Raspberry Pi image processing algorithm must run faster than the frame rate of capturing images. In addition to this, proper synchronization must be ensured.

Here GIL (Global Interpreter Lock) is easy to be used in multithreading in low-level languages compared to python. As python is an interpreted language, the interpreter will not be able to execute code aggressively. This is because the interpreter does not see python script as a whole program. To account for processing speed we must rely on the speed of the interpreter.

In schools we learn that a mental model that is well suited for sequential execution does not match the parallel execution model. Additionally, developing a multi-threaded model becomes more complex very quickly in both developing and debugging compared to single-threaded model.

3.3 Optical Character Recognition Engine

A Tesseract OCR is used in this project in order to read the words in a paragraph accurately. This is an open source engine started as PhD research project at the HP labs, Bristol. Then, HP labs, Bristol and HP scanner division made a joint project in which they showed Tesseract OCR outperformed many other commercial OCR engines. A traditional step-by-step pipeline is required for the processing within the Tesseract OCR.

Connected component analysis is applied in the first step and also the outliners of the component are stored. This step is particularly computational intensive, but also it has a lot of advantages such as, being able to read reversed texts, recognizing the black text on the white background easily. The outliners and the regions are analyzed as blobs

after this stage. With the help of the nested algorithm in OCR for spacing, the text lines are broken into character cells. In the next step, the recognition phase is set to two parts. There is an adaptive classifier to which each word is passed and it recognizes the text more accurately. Also, adaptive classifier helps in deciding whether the font is character or non-character. Therefore adaptive classifier has been suggested for use of OCR engines. A Tesseract OCR engine does not employ a template (static) classifier like most of the other OCR engines.

The adaptive classifier use baseline x-height normalization and other template classifiers find position of the characters based on size normalization. This is the major difference between an adaptive classifier and template classifier. Even though the baseline x-height recognition requires more computational power, it allows for more precise recognition and detection of upper case, lower case characters and digits. If the images are given to the OCR module in the form of clear black texts and white background, the quality of results can be improved. In order to improve speed we have custom pre-processing algorithm. This helps in skipping a default Tesseract OCR step that is, applying Otsu's thresholding method to every single image. The Tesseract delegate option should be set as "self" (tesseract.delegate = self) inorder to disable internal thresholding of the Tesseract OCR.

3.4 Text to speech method

eSpeak is a compact open source software speech synthesizer for English and other languages, for linux and windows which uses a "formant synthesis" method. This allows many languages to be provided in small size. The speech is clear and can be used at high speed, this is the main advantage of using the eSpeak.

To speak text from a file or from stdin and shared library version for use by other programs, eSpeak is available as a command line program (Linux and Windows) on windows this is a DDL. Raspbian OS follows the Advanced Linux Sound Architecture (ALSA) for managing audio devices. Few packages need to be installed to test the sound device through ALSA. They are:

```
apt-get install alsa-utils
apt-get install mpg321
apt-get install lame
```

To test the sound device through ALSA, Install the following packages:

```
modprobe snd-bcm2835
```

These are the few commands that are to be used to install TTS engine (eSpeak) and python module:

```
sudo apt-get install espeak
sudo apt-get install espeak python-espeak
```

This module converts transformed text to audible form. The Raspberry Pi has an on-board Audio Jack,the on-board audio is generated by a PWM output and is minimally filtered. The audio output is produced through the headset connected to the Raspberry Pi.

3.5 Ultrasonic Sensor

High frequency sound waves are generated by ultrasonic sensor. It evaluates the echo which is received back by the sensors. The time interval between sending the signal and receiving the echo is calculated by a sensor to determine the distance to an object. Ultrasonic is like an infrared where it will reflect on a surface in any shape, but ultrasonic has a better range detection compared to infrared.

Ultrasonic Sensor is used to determine distance from obstacles. Ultrasonic sound waves are generated by Ultrasonic sensors which are reflected through obstacles. These reflected signals are then received by the module. Calculating the time delay between transmitted and received signal, distance of obstacle is known from the formula.

s=vt/2; v = velocity of sound = 334 m/s

This distance is compared with threshold distance and warning message is sent to visually impaired people if distance is less than threshold.

Firstly, when the device turns on, the ultrasonic sensor will automatically give the distance measurement of the obstacle in front of the blind, and then the distance measured is stored in the SD card.

3.6 Vibrating Motor

A vibrating motor is essentially a motor that is improperly balanced. Our program triggers the vibrating motor when there is an obstacle in the way. As the obstacle gets closer the intensity of the vibrating motor increases, thereby alerting the user

4. IMPLEMENTATION

The system was built on a Raspberry Pi board running Linux and Python / Open CV Libraries. All compatible versions of Linux on Raspberry Pi including Red Hat, Mandrake, Gentoo and Debian can be used, however, since in this project GPIO pins on Raspberry Pi were extensively used for camera module and USB Wi-Fi dongle, Raspbian

OS was chosen. Raspbian is a Debian based Linux distributed de-facto standard operating system, which comes with pre-installed peripheral units libraries (GPIO, Camera Module). It is jointly maintained by the Raspberry Pi Foundation and community. It also has raspi-config, a menu based tool that makes managing Raspberry Pi configurations much easier than other operating systems, such as setting up SSH, enabling Raspberry Pi camera module etc.

The USB camera captures the image which is in the form of jpeg, png, jpg, bmp etc. This image is then split in the following conditions as described below. These conditions are used for detecting corresponding text and matching it with templates prescribed below:

- Text in the form of Black and White
- Text in the form of colored
- Text with image
- Text and image merged
- Text with different Font Styles

INPUT IMAGE SETS

The analysis of the proposed method is done with the various text input sets (jpeg, png, jpg, bmp etc.). Text in the form of black and white includes a combination of backgrounds, as shown in the figure below. In fig 2(a) the black text is combined with the plain white background. In fig 2(b) the black text has colored background.

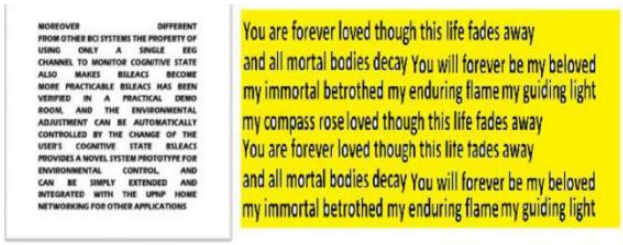

Fig -2: (a) Black and white image (b) colored image

The input includes a text with image and the input with text and images merged which is shown in Fig 3 (a) and (b) respectively.

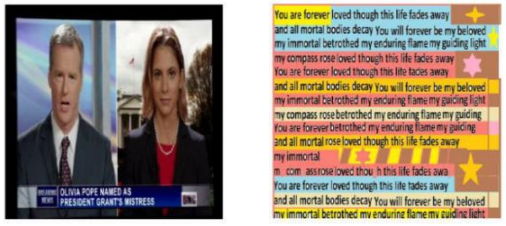

Fig -3: (a) Text with Image (b) Text and image merged

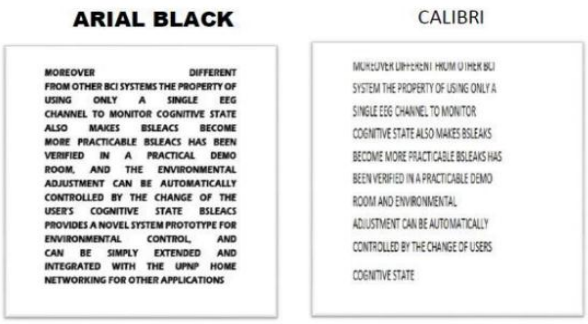

Fig -4: Text with different font styles

Once the image processing of the sample image set is complete, the results can be compared to the actual textual images. Many practical observations were gained while performing data collection and analysis. It was obvious that while the software was robust and capable of detecting text, the corner of the captured text was misinterpreted by Tesseract OCR in some cases. Misinterpretation usually occurs when the speed signs are rotated or the Tesseract OCR converts corner of the frames into letters. After the cropping algorithm applied on pre-processed images, and adding a delay between the frames. The algorithm which has been implemented is very accurate, the total average time for both detection and recognition is ~1.5 frames per second (fps) on a 700 MHz Broadcom chip.

Fig -5: Capturing of image

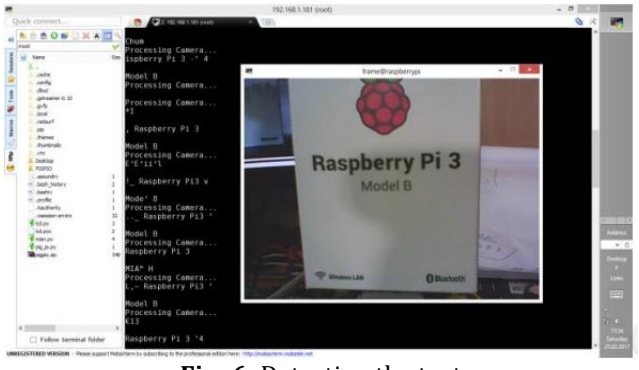

Fig -6: Detecting the text

IMPLEMENTATION USING HARDWARE

The hardware of the proposed work consists of a raspberry pi board interfaced with a USB camera. Wi Fi dongle is connected to the system for internet connection which is taken to Pi through LAN cable. A 5mp camera, a vibrating motor, an ultrasonic sensor is connected to one of the USB port of raspberry pi. A 5V supply is given to Raspberry pi from the system through a power bank.

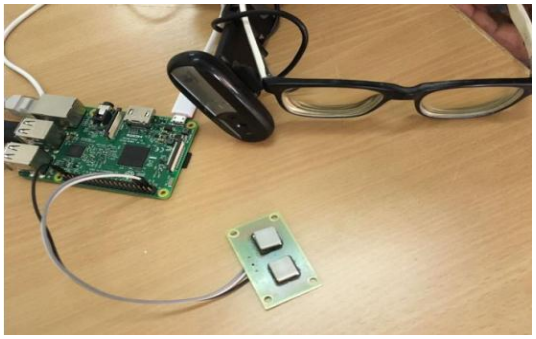

Fig -7: Hardware Implementation

TEST RESULTS

The proposed system is tested with different input sets as, Text in the form of Black and White, Text with different font styles, Text in the form of colored. In all five sets specs inbuiltcamera is tested using OCR process with around 2 samples of 100 letters which yields an accuracy of around 98%.

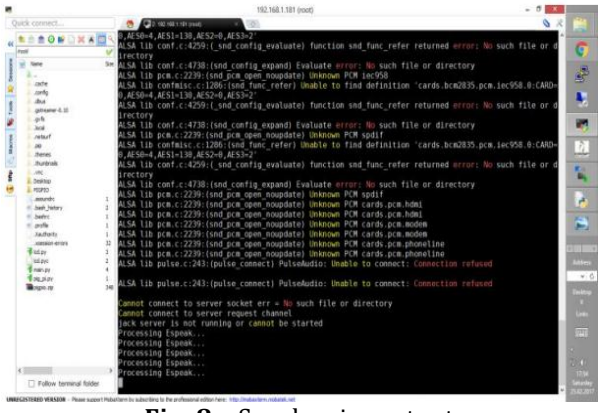

Fig -8: eSpeak voice output

Table -1: Test Results for black and white and color images

Number of letters	100
Words detected	91
Error possibility	9

Table -2: Test Results for texts with different font styles

Number of letters	100
Words detected	94
Error possibility	6

Thus the simulated results of various input sets and the hardware setup of spectacles with different input sets and the recognized output is viewed through Matlab simulator with an audio output.

5. CONCLUSION

The Text to Speech conversion technique has been implemented using Raspberry Pi. The simulation results have been successfully verified and hardware output has been tested for different input samples. The image is processed and read out clearly. This system also makes use of ultrasonic sensors to detect obstacles so that user can be warned by triggering the vibrating motor connected to the raspberry pi. This device is not only efficient but also economical for the visually impaired people.

REFERENCES

[1]. AlexandreTrilla and FrancescAlías. (2013), "SentenceBased Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.

[2]. Alías F. Sevillano X. Socoró J. C Gonzalvo X. (2008), "Towards high-quality next-generation text-to-speech synthesi", IEEE Trans. Audio, Speech, Language Process, Vol. 16, No. 7. pp. 1340-1354.

[3]. Balakrishnan G. Sainarayanan G. Nagarajan R. and Yaacob S. (2007) "Wearable real-time stereo vision for the visually impaired", Vol. 14, No. 2, pp. 6–14.

[4]. Chucai Yi. YingLiTian.AriesArditi. (2014), "Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons", IEEE/ASME Transactions on Mechatronics, Vol. 3, No. 2, pp. 1-10.

[5]. Deepa Jose V. and Sharan R. (2014), "A Novel Model for Speech to Text Conversion", International Refereed Journal of Engineering and Science (IRJES) Vol.3, Issue.1, pp. 39-41.

[6]. Goldreich D. and Kanics I. M. (2003), "Tactile Acuity is Enhanced in Blindness", International Journal of Research And Science, Vol. 23, No. 8,pp. 3439–3445.

[7]. Joao Guerreiro and Daniel Gonçalves (2014), "Text-toSpeech: Evaluating the Perception of Concurrent Speech by Blind People", International journal of computer technology, Vol. 6, No. 8, pp. 1-8.

[8]. J. Liang D. and DoermannH. (2005), "Camera-based analysis of text and documents: a survey", International Journal on Document Analysis and Recognition, Vol.7, No-6, pp. 83-200.

[9]. Manduchi R. and Miesenberger K. (2012), "Mobile Vision as Assistive Technology for the Blind: An Experimental Study", Springer-In Computers Helping People with Special Needs, Vol. 2, No.7383, pp. 9–16.

[10]. Marion A. Hersh Michael A. Johnson (2013), "Assistive Technology for Visually Impaired and Blind", SpringerInternational Journal of Engineering and Technology, Vol. 4, No. 6, pp. 50-69.

[11]. S. Mascaro. And H. H. Asada. (2001) "Finger posture and shear force measurement: Initial experimentation," in Proc.IEEE Int. conf.robot.Autom, Vol. 2,pp. 1857-1862.

[12]. Norman J. F. and Bartholomew A. N. (2011), "Blindness Enhances Tactile Acuity and Haptic 3-D Shape Discrimination", Proceedings of the 4thAugmented Human International Conference, Vol. 73, No. 7, pp. 23–30.

[13]. Pitrelli J. and Bakis R.(2006), "The IBM expressive textto-speech synthesis system for American English", IEEE Trans. Audio, Speech, Lang. Process, Vol. 14, No. 4, pp. 1099–1108.

[14]. PriyankaBacche. ApurvaBakshi, KarishmaGhiya. PriyankaGujar. (2014), "Tech –NETRA (New Eyes to Read Artifact)", International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 3, Issue. 3, pp. 482-485.

[15]. Rissanen, M. J.FernandoS.Pang .N. (2013), "Natural and Socially Acceptable Interaction Techniques for Ringterfaces: Finger-ring Shaped User Interfaces", Springer - In Distributed Ambient and Pervasive Interactions, Vol. 19, No. 6, pp. 52-61.

[16]. RohitRanchal .YirenGuo. Keith Bain and Paul Robinson J (2013), "Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom", IEEE Transactions On Learning Technologies, Vol. 6, No.4, pp. 12-17.

[17]. Rubesh Kumar and Purnima (2014), "Assistive System for Product Label Detection with Voice Output for Blind Users", International Journal of Research in Engineering & Advanced Technology, Vol. 1, Issue. 6, pp. 30-45.

[18]. [Rupali and Dharmale (2015), "Text Detection and Recognition with Speech Output for Visually Challenged Person", Research Gate- International Journal of Engineering Research and Applications, Vol. 5, Issue. 3, pp.84-87.

[19]. Shen, H. and Coughlan, J. M. (2012), "Towards a realtime system for Finding and Reading Signs for Visually Impaired Users", Springer-International Journal in Computers Helping People with Special Need, Vol.2 , No. 1, pp. 41-47.

[20]. Shinnosuke and Takamichi (2014),"Parameter Generation Methods With Rich Context Models for HighQuality and Flexible Text-To-Speech Synthesis", IEEE Journal Of Selected Topics In Signal Processing, Vol. 8, Issue. 2, pp. 239-250 .

[21]. Tapas Kumar Patra and Biplab (2014),"Text to Speech Conversion with Phonematic Concatenation", International Journal of Electronics Communication and Computer Technology (IJECCT) Vol. 2, Issue. 5. pp.223-226.

[22]. Xiang Peng. Fang Liu.Tianjiang Wang.Songfeng Lu (2008), "A Density-base Approach for Text Extraction in Images", Research Gate- International Journal of Communication and Engineering Vol. 8, No. 4, pp. 239-248.

[23]. Bijay Sapkota, Mayank K. Gurung, Prabhat Mali, Rabin Gupta, Er. Manoj Gyawali,(2019) "Virtual eye for Visually Impaired", KEC Conference.